Multiplexed Illumination for Classifying Visually Similar Objects

TAIHUA WANG[1] AND DONALD G. DANSEREAU[2,*]

[1]*Dept of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, NSW, Australia*
[2]*Sydney Institute for Robotics and Intelligent Systems, University of Sydney, NSW, Australia*
[*]*donald.dansereau@sydney.edu.au*

Abstract: We propose the use of multiplexed illumination to enable classification of visually similar objects that cannot normally be distinguished. We construct a compact RGB-IR light stage and develop a method for jointly selecting informative illumination patterns and training a classifier that uses the resulting images. We use the light stage to model training samples and synthesize novel, noise-accurate images that drive the training process and a time-efficient greedy pattern selection scheme. The system delivers fast and accurate classification of previously indistinguishable samples, outperforming fixed-illuminant and conventional noise-optimal patterns. This work has potential applications spanning forgery detection and quality control in agriculture and manufacturing.

1. Introduction

While image classification is a well explored area, the classification of visually similar objects remains challenging. Forged and authentic paintings and currency, diseased and healthy plants, and benign and malignant skin lesions can all be so visually similar as to evade classification with even the most sophisticated modern techniques. In this work we propose the use of a multiplexed active illumination source to facilitate the classification of visually similar objects, and demonstrate it distinguishing challenging objects like the real and synthetic fruit depicted in Fig. 1. Note that the gross differences of colour, size, and shape make it easy to visually distinguish between the different types of fruit, for both humans and standard machine learning algorithms. However, distinguishing synthetic and real samples is much more difficult as the variety of appearance within a class is at least as significant as the differences between classes.

Fig. 1. Visually similar objects, like real and synthetic fruit, fool even the most sophisticated image classifiers. We develop an active multiplexed illumination scheme that allows conventional classifiers to distinguish these challenging examples, and propose a method for selecting illumination patterns that yields fast and accurate classification results.

The idea of extracting information from images by controlling illumination conditions is not new. Prior works controlling the incident light spectrum have distinguished everything from types of alcoholic beverage to forged paintings [1, 2]. Varying the positions of illuminants is also well established: light stage technology spatially distributes and multiplexes light sources to measure visual appearance under a variety of conditions, delivering photo-real relightable models [3].

In this work, we combine the ideas of spectral and spatial multiplexing to improve classification of visually similar objects. While prior work has combined spectral and spatial multiplexing to classify materials [1, 4, 5], our work differs in that it is concerned with object, not material classification. This allows us to distinguish complex non-homogeneous specimens like bills, paintings, lesions, and fruit, opening a new range of applications where visual classification previously failed.

To this end, we construct a compact light stage featuring red, green, blue and near-infrared (NIR) colour channels, each repeated over eight illumination sites. As in other multiplexed illumination schemes, the problem of selecting illumination patterns arises. Signal-to-noise ratio (SNR)-optimal codes are often assumed to maximize system-level performance in such scenarios [6]. However, in this work we show that multiplexing codes can be selected to deliver greater classification accuracy and speed than are possible with SNR-optimal codes.

For code selection, we propose a methodology that uses the light stage in two ways: first we model and synthetically relight training samples to allow joint pattern selection and classifier training in simulation. Then we use the trained patterns and classifier to quickly classify previously unseen samples.

Our key contributions are:

- We propose a multiplexed illumination scheme that allows visually similar objects to be distinguished by conventional classifiers,
- We use light stage capture and noise synthesis to simulate realistic images under different lighting patterns, allowing the training of multiplexing patterns and classifiers in simulation,
- We propose a time-efficient greedy optimization scheme for selecting illumination patterns, and
- We demonstrate our method outperforming naïve fixed-illumination and SNR-optimal multiplexing codes in terms of classification accuracy and speed.

We are also releasing a dataset[1] containing 16000 images of real and synthetic examples of each type of fruit depicted in Fig. 1. Images were collected under a variety of illumination patterns including single and multiple-illuminant states. Because it allows the evaluation of novel multiplexing schemes, we believe the dataset and methodology described here might form the basis for extensive follow-on work.

2. Related work

This work combines the concepts of coded (multiplexed) illumination, multispectral illumination, and image relighting to improve image classification. While prior work has established the utility of these techniques for applications in computer graphics, BRDF acquisition, and material classification, none to our knowledge has yet combined them for improving classification of visually similar objects.

Light multiplexing for enhanced image quality [7] has shown that N single-illuminant images can be obtained at higher SNR by capturing N coded illumination states, followed by demultiplexing. The work establishes multiplexing codes that yield optimal SNR, with follow-on works refining treatment of effects like Poisson noise and saturation [8]. These works underpin much of our understanding of computational imaging [6], which can be seen as delivering improved image quality through multiplexed light capture. A key contribution of our work is to

[1]Code and dataset are available at https://roboticimaging.org/Projects/LSClassifier/

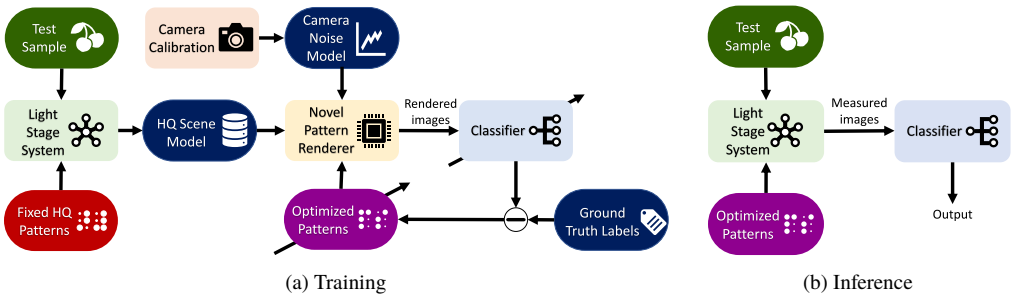


Fig. 2. (a) We form high-quality models of training objects using the compact light stage, and use these in a physically accurate rendering pipeline to simulate different illumination conditions. This allows us to jointly train a classifier and optimize illumination patterns in simulation; (b) At inference time only the optimized illumination patterns are used, yielding fast and accurate classification.

apply multiplexing to improve image classification. Like other works considering applications beyond intensity estimation [9], we show that SNR-optimal codes do not always yield the best overall system performance.

Light stage capture and relighting [10–12] employs multiplexed illumination across a diversity of illuminant poses to capture high-fidelity relightable visual models. Modern examples can produce photo-real relit images of very complex subjects including human faces [3]. In this work we adopt the light stage in two ways: we build a relightable model of the training samples and use this capability to select multiplexed illumination patterns for image classification. Then we apply the resulting patterns using the light stage for fast inference on previously unseen samples.

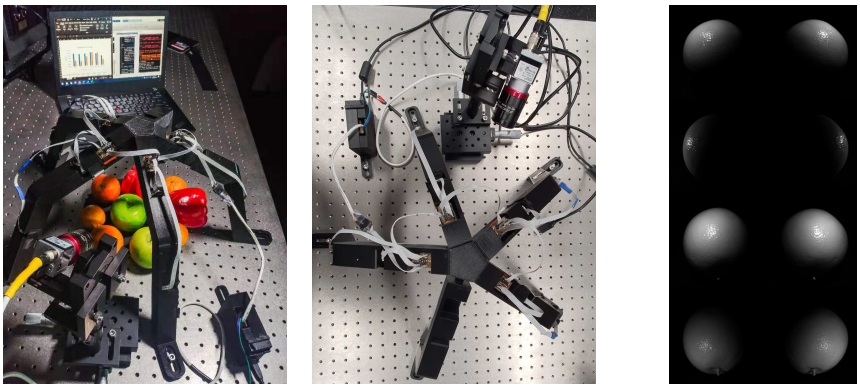
Applying spectral control to improve material classification is also well established [13, 14]. In this work we include an NIR channel to facilitate classification of real and synthetic fruit, noting that its utility in imaging plants has been well established in prior work [15–20]. Including an NIR channel also demonstrates image capture beyond what conventional cameras can achieve, and establishes the viability of the proposed approach for multi-spectral operation.

Prior applications of multiplexed multispectral illumination have generally dealt with rendering [21] or material classification [1, 4]. The later employs coded illumination to carry out single-image pixel-level material classification, and [5] use a light stage for discriminating materials using a few images. Our focus is on classifying objects rather than materials, opening new domains involving heterogeneous, complex objects.

This is the first work to our knowledge that combines the concepts of multispectral and multiplexed illumination to enable object-level classification of very challenging samples. We develop a time-efficient means of deriving multiplexing codes that leverages light stage relighting to carry out the process in simulation. We then employ the light stage at inference time to quickly capture multiple images that allow us to accurately classify objects that would otherwise be visually indistinguishable.

3. Modelling, Training, and Inference

The system proposed in this paper consist of two stages as depicted in Fig. 2. In training, we build high-quality relightable models of our test samples, using the custom-built light stage depicted in Fig. 3. We use these models to select fast and informative multiplexing patterns and simultaneously train a classifier. At inference time, test samples are measured using only the optimized patterns, yielding fast, accurate classification results.



(a) (b) (c)

Fig. 3. The compact light stage shown (a) in profile and (b) from above; (c) depicts typical images obtained with a single illumination site active at a time.

3.1. Scene Model Acquisition

We begin by collecting relightable models of a variety of samples, each under a variety of poses. Both the positions and orientations of the objects vary across the dataset. To build the models we turn the light stage to the task it was originally designed for: multiplexing illumination to build relightable models [12]. Because we want to synthesize, in simulation, noise representative of inference-time conditions, it is important for these models to be captured at a high SNR. We thus operate the camera using long exposures, and take 20 exposures of each sample / pose which we average together, yielding a reduction in noise of about 6.5 dB. We did not measure the final SNR of the high-qualitymodel images, but it is sufficiently high that the synthetic noise added in simulation dominates over the small amount of residual noise present. Note that the examples in this paper were captured with a black background, but the system will work without modification with any low-texture background. Textured backgrounds may require an additional segmentation step, though the classifier should learn to ignore background detail.

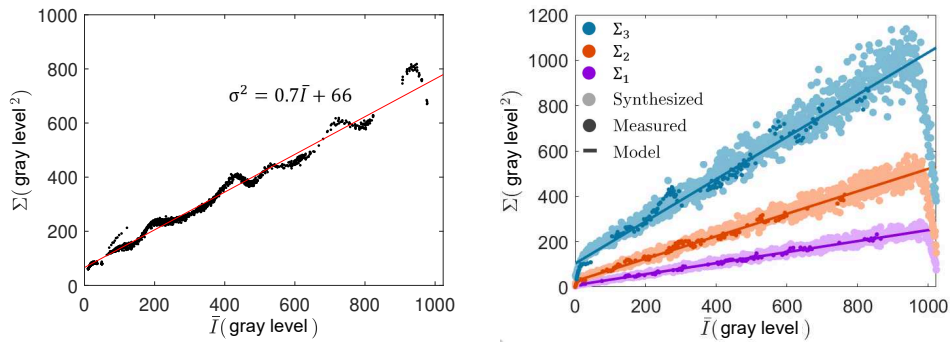
The model acquisition process is much more time-consuming than inference-time image capture, but is worthwhile because it enables fast simulation of a vast variety of illumination states. The total time required to capture high-quality scene models and then simulate thousands of fast illumination patterns is much less than the time it would take to directly test those patterns.

3.2. Camera Noise Characterization

To render physically accurate relit images we require a model of the camera's noise parameters, allowing us to synthetically introduce realistic noise into the rendered images. We characterize the camera when it is configured for inference, including the optics, aperture, and focus, as well as exposure time and gain settings. We then generalize the resulting calibrated camera noise model to other camera settings, allowing us to evaluate, in simulation, a variety of additional imaging scenarios.

Because in this application we are evaluating illumination patterns and test samples that yield a variety of image intensities, it is especially important that our noise model account for intensity-varying Poisson-distributed photon noise. We thus adopt the affine noise model [7], that describes the total noise at a pixel as

$$\sigma^2 = \sigma_p^2 \bar{I} + \sigma_r^2, \tag{1}$$



(a) Characterizing camera noise

(b) Verifying noise model and synthesis

Fig. 4. Characterizing imaging noise: the Basler acA1920-150um at 15 dB gain and 30 ms exposure time yields a characteristic (a) that is well described by an affine noise model (1); (b) In Sec. 4.1 we generalize this model following (2) to three noise characteristics $\Sigma_{1..3}$ corresponding to typical camera gain and exposure settings (8)..(10). The measured (dark) points conform closely to the predicted noise curves (solid lines). We validate noise synthesis by rendering scenes for each camera setting, yielding the lighter points that agree well with the models.

where σ_r^2 is the variance of the fixed read noise, σ_p^2 is the scaling constant for the variance of the intensity-dependent photon noise, and $\bar{I}$ is the mean intensity of each pixel.

There are well established methods for calibrating camera noise [7–9, 22]. Because we have a controlled illumination source in the form of a compact light stage, we calibrate our model by capturing images of a test pattern with smoothly varying reflectivity over a variety of illumination levels. We then find pixel means and variances over multiple exposures, and fit the affine noise model to the resulting observations. An example characterization for the camera used in this work is shown in Fig. 4. Note that this characterization will not hold for very long exposure times, for which dark current would need to be included. However, we employ the characterization only for fast acquisition times, and can therefore neglect dark current.

We generalize the noise model to different camera settings by considering the ratio of a new gain G and exposure time E to their calibrated levels G_0, E_0. The total variance for the generalized noise model becomes

$$\Sigma = \frac{G^2}{G_0^2}\left(\frac{E}{E_0}\sigma_p^2\bar{I} + \sigma_r^2\right). \tag{2}$$

Note that while photon noise varies with both gain and exposure time, read noise is well approximated as varying only with gain. Three generalized noise characteristics are shown in Fig. 4b – note that pixel value saturation limits noise variance for bright pixels, and this manifests as a decrease in variance at the upper end of the range.

3.3. Simulating Novel Illumination Patterns

For a light stage with N illumination sites, we describe an illumination state with a column vector $\boldsymbol{w}$ of length N, where $w_i \in [0, 1]$. Each relightable model contains as many images as there are illuminants, N. For images with a total of N_{pix} pixels, we reshape each into a column vector and stack them into a model L of size $N_{pix} \times N$. Then, rendering a noise-free image can be written as

$$\bar{I} = L\mathbf{w}, \tag{3}$$

where $\bar{I}$ indicates the mean image intensity at each pixel.

To simulate noise at a prescribed camera gain and exposure, we use the generalized noise model (2) to yield a per-pixel physically derived variance Σ. For each pixel we then draw from a Gaussian distribution

$$I \sim \mathcal{N}(\bar{I}, \Sigma), \tag{4}$$

clipping the resulting value to lie within the gray level range of the camera.

3.4. Joint Pattern Selection and Classifier Training

The ability to simulate illumination states with physically realistic noise allows us to train and evaluate classifiers under a variety of novel conditions. In particular, we can collect a sequence of M images, each corresponding to a different illumination state $\boldsymbol{w}$. Taken together, these acquisition states form an $N \times M$ illumination matrix W. Gathering more images takes more time, but can increase classifier performance. The goal of the training process is thus to select the illumination matrix W and classifier that together maximize classifier accuracy and speed.

Fig. 2 shows how rendered images drive the joint optimization of the illumination matrix and the classifier. In the main training loop, a set of illumination patterns drives the rendering of a new set of images. These are used to train and evaluate a classifier, and this process repeats, optimizing towards improved classifier performance by selecting better illumination matrices.

We anticipate a variety of approaches could be adopted for this optimization process. In this work we propose a greedy scheme, depicted in Fig. 5, that incrementally grows a multiplexing matrix one column at a time. This begins by determining the best-performing single-image illumination pattern. It then builds on this by incrementally adding images, selecting the best-performing illumination pattern for each. At each step of the process illumination patterns are evaluated by training and evaluating a classifier and measuring mean accuracy, as depicted in Fig. 5. The process ends when no illumination pattern improves performance of the system. The practitioner can then choose to use the whole illumination matrix, or to use a subset to obtain faster but less accurate results.

For our implementation, we evaluate individual patterns exhaustively. That is, for N illumination sites, we test 2^N illumination states. For an M-pattern sequence, we test a total of $M2^N$ illumination matrices. Each pattern is evaluated by rendering relit versions of all available training scenes, and splitting the result to train and test the classifier. To make maximal use of the available models, we repeat this process for different random selections of training and test data. We also re-render images in each pass, to avoid overfitting to a particular instance of randomly selected noise.

We employ a multi-class support vector machine (SVM)-based classifier. While convolutional neural networks (CNNs) are powerful their behaviour is also more complex than SVMs, and this might obfuscate important features of the multiplexing and pattern selection processes. SVM behaviour is well understood and is directly revealing of system-level phenomena. We expect this work to generalize well to CNN-based classifiers and leave this as future work.

To drive the SVM, we concatenate descriptors computed on regular grids on each of the M differently illuminated images. We do not employ a bag of words (BoW) as these cluster and equate features with similar appearance, an undesirable trait when classifying visually similar classes. Rather we rely on a diversity of sample poses to provide rotational and translational invariance, and the consistency of the light stage camera pose to limit apparent scale variations.

For a given illumination matrix W, different strategies can be used for controlling camera exposure and gain. We fix the camera exposure to a desired setting that will ultimately determine the acquisition and thus inference speed for the system. For each matrix under consideration, we select a single camera gain appropriate for the brightest image in the set, yielding a single-setting acquisition process.

Inference is carried out on previously unseen samples using the optimized illumination matrix and classifier. As depicted in Fig. 2, the illumination patterns drive the light stage to quickly

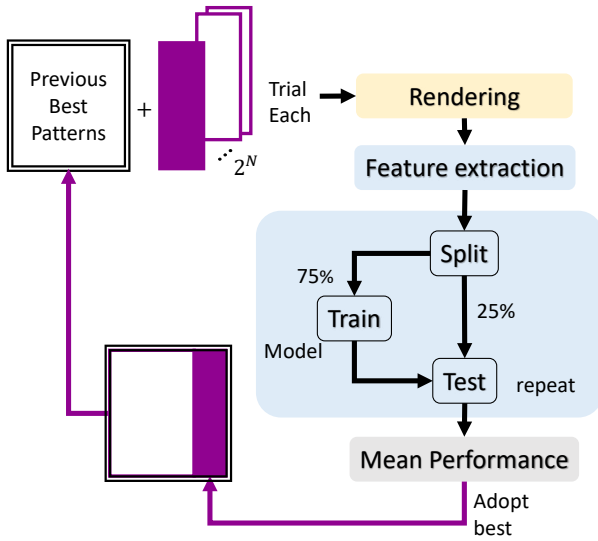


Fig. 5. Greedy pattern selection: the illumination matrix (top left) is incrementally grown one pattern at a time. For each new column, we perform an exhaustive search over all 2^N possible new patterns, for each rendering the corresponding relit imagery, training and testing a classifier, and selecting the one that yields the highest overall accuracy. Train and test are repeated to make best use of the dataset.

capture a set of images which are passed to the classifier. The inference process can be much faster than the modelling process, because the illumination patterns are trained to operate with a fast camera exposure time, and are selected to yield accurate results with only a few images.

3.5. SNR-Optimal Illumination Patterns

As an alternative to the greedy pattern selection scheme described above, we also present a method based on the theory of multiplexed illumination, in which the goal is to estimate a set of single-illuminant images from a set of coded multiple-illuminant images [7]. Multiplexing illumination in this way boosts the SNR of the recovered images, and SNR-optimal multiplexing is generally held to yield maximal performance in computational imaging systems. However, as we will show, this method does not adapt illumination patterns to the problem space, yielding lower performance compared with our proposed greedy scheme.

To correctly account for a specific calibrated camera noise characteristic including Poisson noise, we adopt an optimization-based approach that adjusts the multiplexing matrix W to maximize peak signal-to-noise ratio (PSNR), assuming an average scene reflectance $\bar{R}$ [6]. Starting with the identity matrix, we employ a stochastic optimization scheme in which each iteration perturbs the previous best matrix and evaluates the result. Checking the condition number allows us to reject non-invertible matrices.

To evaluate a matrix W we estimate the noise variance expected in the measured imagery by generalising (1) as

$$\Sigma_W = \mathrm{diag}\left(\sigma_p^2 \bar{R} \sum_{col} W + \sigma_r^2\right), \tag{5}$$

where the summation over columns of W estimates the overall brightness for each illumination state, and the average scene reflectance $\bar{R}$ converts this to an average pixel intensity. We then

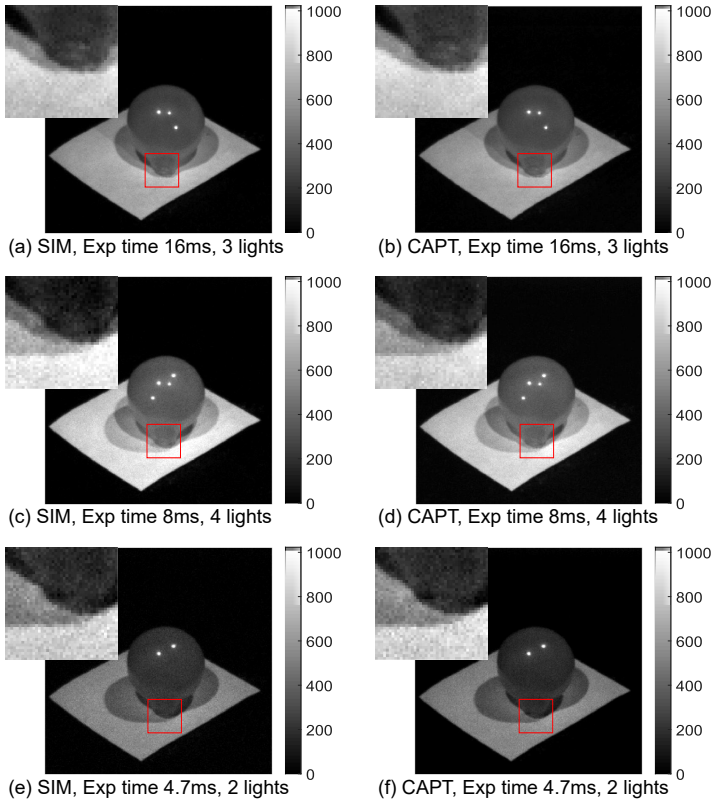


Fig. 6. We validate rendering by generating imagery at different illumination conditions and camera settings (left), and comparing with the corresponding captured imagery (right). Both the change in appearance due to different lighting patterns and the appearance of noise are a very close match.

estimate the mean squared error (MSE) of the recovered single-illuminant images following [9]

$$\mathrm{MSE} = \frac{1}{n} \operatorname{tr}\left[(W^T \Sigma_W{}^{-1} W)^{-1} \right]. \tag{6}$$

Driving the estimated MSE to a minimum yields the PSNR-optimal multiplexing matrix.

Note that this SNR-optimal approach always yields a square illumination matrix and does not obviously provide for early termination in which subsets of illumination states offer superior performance.

4. Results

To evaluate our methodology we constructed a compact light stage prototype as depicted in Fig. 3. The prototype features a five-leg design, with a 223 mm radius and 240 mm height. Its eight illumination sites are distributed across four of its legs, four mounted on upper leg segments, and four mounted on lower.

Each illumination site has four LEDs centered on Red (615 nm), Green (515 nm), Blue (460 nm), and NIR (850 nm) colour bands. A diffuser on each unit softens specular highlights and reduces apparent position variation with colour. LED control is via an extensible daisy-chained circuit.

We employ a Basler acA1920-150um monochrome machine vision camera with an Edmund Optics NIR-VIZ 6mm infrared-compatible lens. The 2.3 megapixel sensor measures 9.2×5.8 mm, with 4.8 micron pixels. We crop the captured images to correspond to the working area, yielding square images 800 pixels on side. The camera and LEDs are hardware-triggered by a microcontroller to maintain synchronization.

We evaluate our system using the five types of synthetic and real fruit depicted in Fig. 1. The two sets of experiments in this paper employ different sample sets, and evaluation on captured imagery uses samples not seen during training. A high quality relightable model of each training sample was collected over 20 different poses, as described in Sec. 3.1. We employed a single-illumination approach to model acquisition, though this could be accelerated using multiplexing as in [7]. In total we collected 6400 images for training, and another 12800 for testing, 6400 for each of the proposed and SNR-optimal approaches.

4.1. Validating Rendering and Noise Synthesis

Here we validate the scene model acquisition, camera noise characterization, and physically realistic rendering processes described in Sects. 3.1, 3.2, and 3.3,

Camera calibration proceeds by imaging a monochromatic object under increasing illumination intensities. We characterize the camera at 15 dB gain and 30ms exposure time, collecting 120 images at each of a range of illumination intensities, yielding the characteristic shown in Fig. 4. Following (1), a line fit yields the affine noise model

$$\sigma^2 = 0.7\bar{I} + 66. \tag{7}$$

We wish to generalize the camera's characterization to other gain and exposure settings. We thus select three typical exposure times of 84, 42, and 22.5 ms, and pair each with a corresponding gain setting of 6, 12, and 17.5 dB, with the latter approaching the camera's maximum gain. We label each of these three settings $\Sigma_{1..3}$, and use (2) to estimate the camera's noise characteristic at each:

$$\Sigma_1 : \sigma_p^2 = 0.25,\ \sigma_r^2 = 8.35, \tag{8}$$

$$\Sigma_2 : \sigma_p^2 = 0.50,\ \sigma_r^2 = 33.23, \text{and} \tag{9}$$

$$\Sigma_3 : \sigma_p^2 = 0.94,\ \sigma_r^2 = 117.37. \tag{10}$$

We validate each of these new noise models by characterizing the camera at each setting $\Sigma_{1..3}$. The results, seen in Fig. 4, show good agreement between each model (solid line) and measured characteristic (dark points). A slight difference in the read noise is evident with the noisier setting Σ_3, but the difference is within a few gray levels and the overall fit is good.

We also validate the noise introduced in the rendering pipeline by simulating images at each of the camera settings $\Sigma_{1..3}$. Repeating each rendering and taking statistics over the set allows us to plot noise characteristics for the rendered images, seen as faint points in Fig. 4, and showing good agreement with the targeted characteristic in each case. Note the decrease in variance near the top end of the curves is due to saturation.

To further validate the rendering process, we synthesized and captured images across several illumination states and camera settings. Typical examples are shown in Fig. 6, showing good agreement between rendered and captured images. As exposure time decreases, top to bottom, we perceive a clear decrease in SNR, as expected.

4.2. Pattern and Classifier Training

Our light stage has eight illumination sites, each having four LEDs. We could optimize patterns over all combinations of the 32 LEDs. Although we expect this would yield strong results, it also

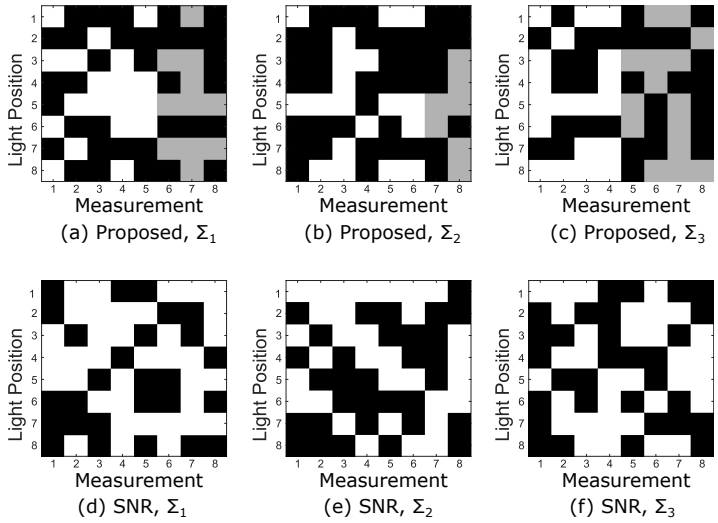


Fig. 7. Multiplexing patterns optimized for the three noise characteristics $\Sigma_{1..3}$, with white corresponding to a light being on: (top) patterns optimized with the proposed greedy method, and (bottom) SNR-optimal patterns. The proposed method supports early termination of training when performance peaks – shaded columns indicate patterns that did not improve performance.

represents a massive search space, even for the efficient greedy method proposed here. We thus proceed by optimizing patterns over the eight illumination sites, repeating the spatial pattern across the four colour channels. This results in a total of 256 possible illumination states for each image in the sequence. Our results show that this reduced search space is sufficient for learning to distinguish very similar objects.

To drive the classifier we concatenate features as described in Sec. 3.4, employing a histogram of oriented gradients (HOG) feature on a regular grid with cell size 12 and block size 10. We forego the use of a BoW as this may be incompatible with classification of visually similar objects, and instead emphasize pose diversity in our training data, and downscale all images to 120×120 pixels to keep the feature space tractable.

For each illumination matrix evaluated as part of the optimization loop, the classifier training and evaluation are repeated 400 times. During training the rendered and feature-extracted imagery is randomly split between 75% training and 25% validation data. Repeating the test makes the most of the dataset, and the average performance is used to drive the optimization process. As discussed in Sec. 4 we used 6400 images during this training/validation process, and 6400 different images for testing each of the SNR-optimal and proposed methods.

Carrying out the proposed greedy optimization scheme for each camera setting $\Sigma_{1..3}$ yielded the illumination matrices depicted in Fig. 7. As described in Sec. 3.3, each column of each matrix corresponds to a different measurement, i.e. a different illumination state, and each row indicates the state of a specific illumination site, on or off.

The greedy optimisation scheme is capable of early termination, in which training halts once peak performance is reached. However, we completed training across eight patterns for all methods, to provide a more complete evaluation. Patterns that did not result in an increase in performance are grayed out in the figure. Also shown are the SNR-optimal patterns selected for the same camera settings, following the method in Sec. 3.5.

4.3. Classification Accuracy

We evaluated the trained illumination patterns shown in Fig. 7 classifying synthetic and captured images, with results shown in Fig. 8 and Table 1. Here accuracy is taken as the number of correct classifications over the total number of classifications, and is shown in Table 1 both for each type of fruit, and overall. The table also depicts precision, recall and F1 score, found as the macro average over all classes following the standard definitions for each of these quantities.

Evaluating on simulated imagery involved using the same noise-accurate rendering pipeline used in training to synthesize new images at each of the three camera settings $\Sigma_{1..3}$. Evaluation on captured imagery involved driving the light stage with the trained illumination patterns to capture previously unseen samples. Noting the similarity between simulation results across settings, we evaluated the captured imagery at only the most challenging setting, Σ_3.

We evaluated each pattern incrementally, testing first with a single captured image per sample, corresponding to the first column of each pattern, then adding a second image, and so on. For the proposed method the required classifier for each image count was already available from the pattern selection process. To evaluate the SNR-optimal patterns, a classifier was trained for each image count using noise-accurate rendered imagery.

As seen in Fig. 8, tests on simulated imagery at all three noise levels $\Sigma_{1..3}$ showed the proposed method offering a strong performance advantage over the SNR-optimal patterns. The proposed method consistently showed higher accuracy, and grew in accuracy more quickly with image count, reaching peak performance at four or five images, while the SNR-optimal method peaked at seven or eight images. Similarity of results across the three noise levels imply the system might be pushed to faster operation.

For captured imagery, the two methods performed more similarly, with the proposed showing slightly improved performance, especially at low image counts. Importantly, our method was able to perform well despite operating on previously unseen samples captured at the fastest acquisition rate. Both methods performed better than the simulation experiments, and we hypothesize the discrepancy in performance is due to the captured-imagery samples being less challenging than the samples evaluated in simulation.

Table 1 shows a comparison of performance for rendered and captured test imagery at the camera setting Σ_3. Here we combine synthetic and real samples to report average performance for each of the five types of fruit. Results are shown for the SNR-optimal and proposed patterns, as well as for two naïve methods. All methods employed the same classifier, but the naïve approaches did not use multiplexed illumination, instead collecting a single image with all LEDs active to yield maximal illumination and signal. The naïve approach was evaluated both with and without a BoW.

The two multiplexing-based methods were evaluated using their optimal image counts. The proposed method naturally provides for this form of shortcut evaluation, and although the SNR-optimal training method does not, the evaluation in Fig. 8 allowed us to select the appropriate image count to give this method the best results.

The results show both multiplexing methods significantly outperforming the naïve approaches, and the proposed method showing higher overall accuracy, precision, recall, and F1 score than the SNR-optimal approach. Training and inference times, reported in Table 2, show the proposed method also allows for faster operation compared with SNR-optimal method. For the SNR-optimal approach we show timing results for the full illumination matrix, as there is no obvious way to perform shortcut evaluation with this method to consistently yield faster performance. Overall, the proposed method shows the highest accuracy and fastest performance.

5. Conclusions and Future Work

In this paper we presented an active illumination-based means of making a very challenging visual classification task achievable. We applied light stage technology in two ways: driving

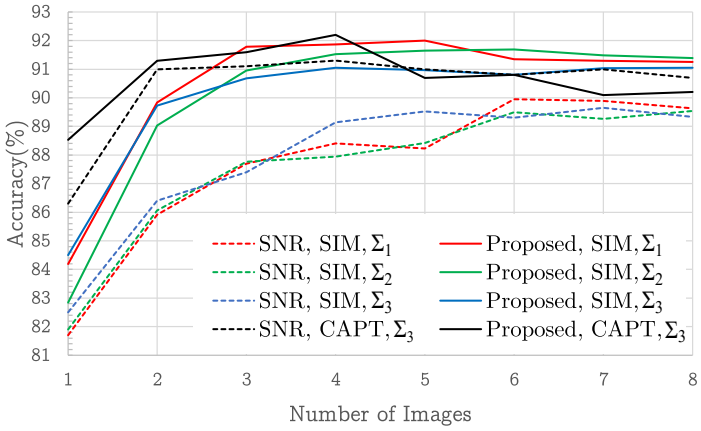


Fig. 8. Two experiments on different samples: "SIM" uses noise-accurate relit images for both training and testing, while "CAPT" is evaluated on previously unseen samples captured using trained illumination patterns. Accuracy vs. image count shows the proposed (solid) method outperforming the SNR-optimal (dashed) approach across all three noise characteristics $\Sigma_{1..3}$ with synthetic imagery, and similar performance on captured imagery. The proposed approach performs better with fewer images, allowing faster performance for a given accuracy.

Table 1. Peak performance for proposed, SNR-optimal, and a naïve fixed-illumination approaches, tested on simulated and captured data. The naïve approaches show weak performance that is dramatically improved upon by both the SNR-optimal and proposed illumination schemes. The proposed method shows the highest overall performance.

	Accuracy (%)								
Method	G.Apple	R.Apple	Grape	Banana	Orange	Overall	Prec.	Recall	F_1
Naïve SIM	30.00	66.00	69.00	58.00	72.00	59.00	57.95	59.00	57.08
Naïve BoW SIM	53.00	47.00	97.00	68.00	63.00	65.60	65.21	65.60	64.25
SNR SIM	89.25	91.70	**96.65**	71.90	**99.55**	89.81	89.81	90.19	90.00
Proposed SIM	**91.65**	**93.50**	96.15	**74.50**	99.45	**91.05**	**91.04**	**91.22**	**91.13**
SNR CAPT	84.15	**95.00**	**98.50**	79.75	**99.25**	91.33	91.33	91.45	91.39
Proposed CAPT	**94.20**	90.70	96.85	**83.45**	95.80	**92.20**	**92.20**	**92.40**	**92.30**

inference-time illumination patterns to allow fast classification of visually similar objects, and accurately rendering relit samples to drive pattern selection and classifier training.

We constructed a compact RGB-IR light stage and used it to classify visually similar objects. We compared against naïve approaches, establishing that without the light stage the classification task is not well achieved. We then proposed two alternative illumination pattern selection schemes, one that optimizes SNR in the conventional multiplexing sense [7], and one greedy optimization scheme that incrementally grows a pattern while training a matching classifier. We showed the greedy approach achieves better overall performance on simulated and captured test samples, and performs better with fewer images than the SNR-optimal pattern.

Although specialized illumination hardware is required to carry out our proposed training and classification scheme, this might find application in a range of situations where conventional approaches cannot effectively operate, and that require fast classification of visually similar

Table 2. Training and inference time for proposed and SNR-optimal approaches, tested on simulated and captured data. The proposed method is faster because peak performance is attained with fewer images, and training naturally terminates once performance peaks.

Method	Train (ms)	Infer (ms)
SNR SIM	1197	35
Proposed SIM	**650**	**24**
SNR CAPT	1276	756
Proposed CAPT	**651**	**376**

samples. Examples are in forgery detection, disease detection for crops and plants, diagnosis of dermal lesions and spots, and quality control for manufacturing and agriculture.

We see several avenues for improvement and hope this work will inspire further development. The pattern selection process only considered the eight illumination positions, repeating the pattern over colour channels. Jointly training chromatic and spatial domains yields a larger search space but should also offer faster, more accurate results. Extension to consider more sophisticated classifiers is also possible, and we expect that jointly learning illumination patterns, feature extraction and classification might allow operation on even more challenging cases. Extension to dynamic scenes may also be possible by compensating for or explicitly learning to tolerate motion.

Acknowledgments

We would like to thank the University of Sydney Aerospace, Mechanical and Mechatronic Engineering FabLab for their support.

Disclosures

The authors declare no conflicts of interest.

References

1. Y. Asano, M. Meguro, C. Wang, A. Lam, Y. Zheng, T. Okabe, and I. Sato, "Coded illumination and imaging for fluorescence based classification," in *European Conference on Computer Vision (ECCV),* (2018), pp. 502–516.
2. T. C. Moran, A. D. Kaye, A. Rao, and F. R. Bueno, "The roles of X rays and other types of electromagnetic radiation in evaluating paintings for forgery and restoration," J. Forensic Radiol. Imaging **5**, 38–46 (2016).
3. K. Guo, P. Lincoln, P. Davidson, J. Busch, X. Yu, M. Whalen, G. Harvey, S. Orts-Escolano, R. Pandey, J. Dourgarian, D. Tang, A. Tkach, A. Kowdle, E. Cooper, M. Dou, S. Fanello, G. Fyffe, C. Rhemann, J. Taylor, P. Debevec, and S. Izadi, "The relightables: Volumetric performance capture of humans with realistic relighting," ACM Transactions on Graph. (TOG) **38**, 1–19 (2019).
4. C. Kampouris and A. Ghosh, "ICL multispectral light stage: building a versatile LED sphere with off-the-shelf components," in *Eurographics Workshop on Material Appearance Modeling,* (Eurographics Association, 2018), pp. 1–4.
5. C. Wang and T. Okabe, "Joint optimization of coded illumination and grayscale conversion for one-shot raw material classification," in *British Machine Vision Conference (BMVC),* (2017).
6. K. Mitra, O. S. Cossairt, and A. Veeraraghavan, "A framework for analysis of computational imaging systems: role of signal prior, sensor noise and multiplexing," Transactions on Pattern Analysis Mach. Intell. **36**, 1909–1921 (2014).
7. Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "Multiplexing for optimal lighting," Transactions on pattern analysis machine intelligence (TPAMI) **29**, 1339–1354 (2007).
8. N. Ratner and Y. Y. Schechner, "Illumination multiplexing within fundamental limits," in *Computer Vision and Pattern Recognition (CVPR),* (IEEE, 2007), pp. 1–8.
9. M. Alterman, Y. Y. Schechner, and A. Weiss, "Multiplexed fluorescence unmixing," in *International Conference on Computational Photography (ICCP),* (IEEE, 2010), pp. 1–8.
10. I. Sato, T. Okabe, Y. Sato, and K. Ikeuchi, "Using extended light sources for modeling object appearance under varying illumination," in *International Conference on Computer Vision (ICCV),* vol. 1 (IEEE, 2005), pp. 325–332.

11. A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec, "Performance relighting and reflectance transformation with time-multiplexed illumination," in *ACM Transactions on Graphics (TOG),* vol. 24 (ACM, 2005), pp. 756–764.
12. P. Debevec, "The light stages and their applications to photoreal digital actors," SIGGRAPH Asia **2** (2012).
13. L. Zhaoxiang and L. Gang, "Apple maturity discrimination and positioning system in an apple harvesting robot," New Zealand J. Agric. Res. **50**, 1103–1113 (2007).
14. H. Blasinski, J. Farrell, and B. Wandell, "Designing illuminant spectral power distributions for surface classification," in *Computer Vision and Pattern Recognition (CVPR),* (2017), pp. 2164–2173.
15. M. Nagata, J. G. Tallada, and T. Kobayashi, "Bruise detection using NIR hyperspectral imaging for strawberry (fragaria× ananassa duch.)," Environ. Control. Biol. **44**, 133–142 (2006).
16. S. D. Roy, D. H. Das, M. K. Bhowmik, and A. K. Ghosh, "Bruise detection in apples using infrared imaging," in *International Conference on Electrical and Computer Engineering (ICECE),* (IEEE, 2016), pp. 118–122.
17. M. Teena, A. Manickavasagan, L. Ravikanth, and D. Jayas, "Near infrared (NIR) hyperspectral imaging to classify fungal infected date fruits," J. Stored Prod. Res. **59**, 306–313 (2014).
18. Q. Lü, M.-J. Tang, J.-R. Cai, J.-W. Zhao, and S. ViTTayapaduNg, "Vis/NIR hyperspectral imaging for detection of hidden bruises on kiwifruits," Czech J. Food Sci. **29**, 595–602 (2011).
19. B. M. Nicolaï, E. Lötze, A. Peirs, N. Scheerlinck, and K. I. Theron, "Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging," Postharvest biology technology **40**, 1–6 (2006).
20. J. P. Wachs, H. I. Stern, T. Burks, and V. Alchanatis, "Low and high-level visual feature-based apple detection from multi-modal images," Precis. Agric. **11**, 717–735 (2010).
21. C. LeGendre, X. Yu, D. Liu, J. Busch, A. Jones, S. Pattanaik, and P. Debevec, "Practical multispectral lighting reproduction," ACM Transactions on Graph. (TOG) **35**, 1–11 (2016).
22. C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in *Computer Vision and Pattern Recognition (CVPR),* vol. 1 (IEEE, 2006), pp. 901–908.